Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity

Immunoglobulin heavy chain (*IGH*) and light chain gene sequences of 27 patients with primary cold agglutinin disease (CAD) were studied to find features explaining the heterogeneity of clinical presentation and disease activity.

CAD is a hemolytic anemia mediated by monoclonal IgM anti-I autoantibodies. CAD represents 15% of all cases of autoimmune hemolytic anemia with an incidence of $1x10^{-6}$ per year. Previously, we demonstrated that CAD is caused by a low-grade clonal B-cell lymphoproliferative disease of the bone marrow with a typical histology that is different from lymphoplasmacytic lymphoma and, accordingly, does not display the *MYD88* L265P mutation.[1] Almost all patients display circulating monoclonal antibodies encoded by the *IGH* gene *V4-34*. The framework region 1 (FR1) of *IGHV4-34* encodes Gln^{6}-Trp^{7} (QW) and Ala^{24}-Val^{25}-Tyr^{26} (AVY) sequences that determine I antigen-binding.[2] *IGHV4-34*, as opposed to other *IGHV4* family genes, has a unique N-glycosylation sequon Asn–X–Ser/Thr located within the complementarity-determining region 2 (CDR2). Inactivation of this sequon by somatic hypermutation (SHM) might increase the accessibility of the antigen-binding pocket.[3] Hitherto, the immunoglobulin (IG) light chain usage in CAD has not been systematically studied.

We studied *IGH* and *IG* light chain gene sequences of 27 CAD patients. Productive *IGHV4-34* gene rearrangements were identified in 81% (22/27) of patients (Table 1 and *Online Supplementary Table S1*). In 4 patients (15%), productive rearrangements could not be identified, whereas a productive *IGHV3-23* gene rearrangement was detected in 1 patient (CAD-5). No significant homology of CDR3 regions was found between *IGHV* sequences. In all but one case (CAD-7), the AVY sequence, required for I antigen-binding, was conserved. In CAD-7, the AVY was mutated to VVY (Figure 1A), replacing alanine (A) with valine (V), that has similar biophysical properties. The N-glycosylation sequence within the *IGH* CDR2 region, also known to affect I antigen-binding, was mutated in 8 patients (36.5%), whereas no mutations were present in 14 patients. However, in 6 of the latter patients mutations in flanking residues were detected (27%) (Figure 1A; Table 1 and *Online Supplementary Table S1*). Mutations flanking the N-glycosylation region are also known to modulate glycosylation efficacy.[3] A mutation hotspot in *IGH* FR3 was identified at the germline amino acid sequence Lys^{90}-Leu^{91}-Ser^{92} (KLS) (Figure 1A). All but 3 patients showed mutations in this region with 7/22 (32%) revealing 1 mutation, 11/22 (50%) 2 mutations and 1/22 (4.5%) 3 mutations (Table 1 and *Online Supplementary Table S1*). Clonal rearrangement of the immunoglobulin κ (*IGK*)*V3-20* gene was detected in 16/27 (59%) of patients, clonal rearrangement of the *IGKV3-15* gene was identified in 4/27 (15%) of patients, whereas other *IG* light chain genes were rearranged in 4/27 (15%) of patients (Table 1 and *Online Supplementary Table S1*). In 3/27 (11%) of patients clonal *IG* light chain gene rearrangements could not be demonstrated. Of interest, 9 of the patients with *IGKV3-20* rearrangement displayed highly homologous CDR3 regions (Figure 2).

Table 1. Analysis of *IGH* and *IG* light chain sequences and correlation between molecular findings and clinical features.

Analysis of *IGH* and *IG* light chain gene sequences	Number of patients (%)
Rearranged *IGH* genes	
Productive *IGHV4-34*	22/27 (81%)
Unproductive/not detected	4/27 (15%)
Productive *IGHV3-23*	1/27 (4%)
Mutations in N-glycosylation sequence in *IGHV4-34* CDR2	
No mutations	8/22 (36,5%)
Mutations in flanking residues only	6/22 (27%)
Mutations in the core region (inactivating mutations)	8/22 (36,5%)
Mutations in Lys-Leu-Ser (KLS) sequence in *IGHV4-34* FR3	
No mutations	3/22 (13.5%)
1 mutation	7/22 (32%)
2 mutations	11/22 (50%)
3 mutations	1/22 (4.5%)
Rearranged *IG* light chain genes	
Clonal *IGKV3-20*	16/27 (59%)
Clonal *IGKV3-15*	4/27 (15%)
Other *IG* light chain genes	4/27 (15%)
Not detected	3/27 (11%)
Amino acid sequence of *IGKV3-20* CDR3 region*	
Highly homologous	9/16 (56%)
Mutated	6/16 (38%)
Correlation between molecular findings and clinical features	**Pearson correlation coefficient (r), *P*-value**
Mutations in *IGHV4-34* CDR2 N-glycosylation site with hemoglobin levels at diagnosis	r= - 0.450 *P*=0.036
Mutations in *IGHV4-34* FR3 KLS sequence with hemoglobin levels at diagnosis	r= - 0.462 *P*=0.030
Mutations in *IGHV4-34* CDR2 N-glycosylation site + FR3 KLS sequence with hemoglobin levels at diagnosis	r= - 0.571 *P*=0.005
Stereotyped *IGK* CDR3 region with lower age at diagnosis	r= 0.554 *P*=0.011
% of mutations in *IGKV3-20* gene with age at diagnosis	r= 0.647 *P*=0.012

**One sample has a suboptimal CDR3 sequence.*

IGHV and *IGKV* genes were variably mutated. *MYD88* L265P mutation, typical for lymphoplasmacytic lymphoma, was not detected in any of the analyzed samples.

We found lower hemoglobin levels at diagnosis to be correlated with inactivating mutations of the N-glycosylation site located within *IGH* CDR2 (*P*=0.036) (Table 1). The N-glycosylation site is located within the antigen-binding pocket (Figure 1B). It has previously been suggested that an intact N-glycosylation site diminishes specific binding of the antigen-binding site of the antibody.[3] Thus, the observation that patients with mutations at the N-glycosylation site had more severe anemia is consistent with I antigen-binding by the antibody antigen-binding site, as previously demonstrated by functional

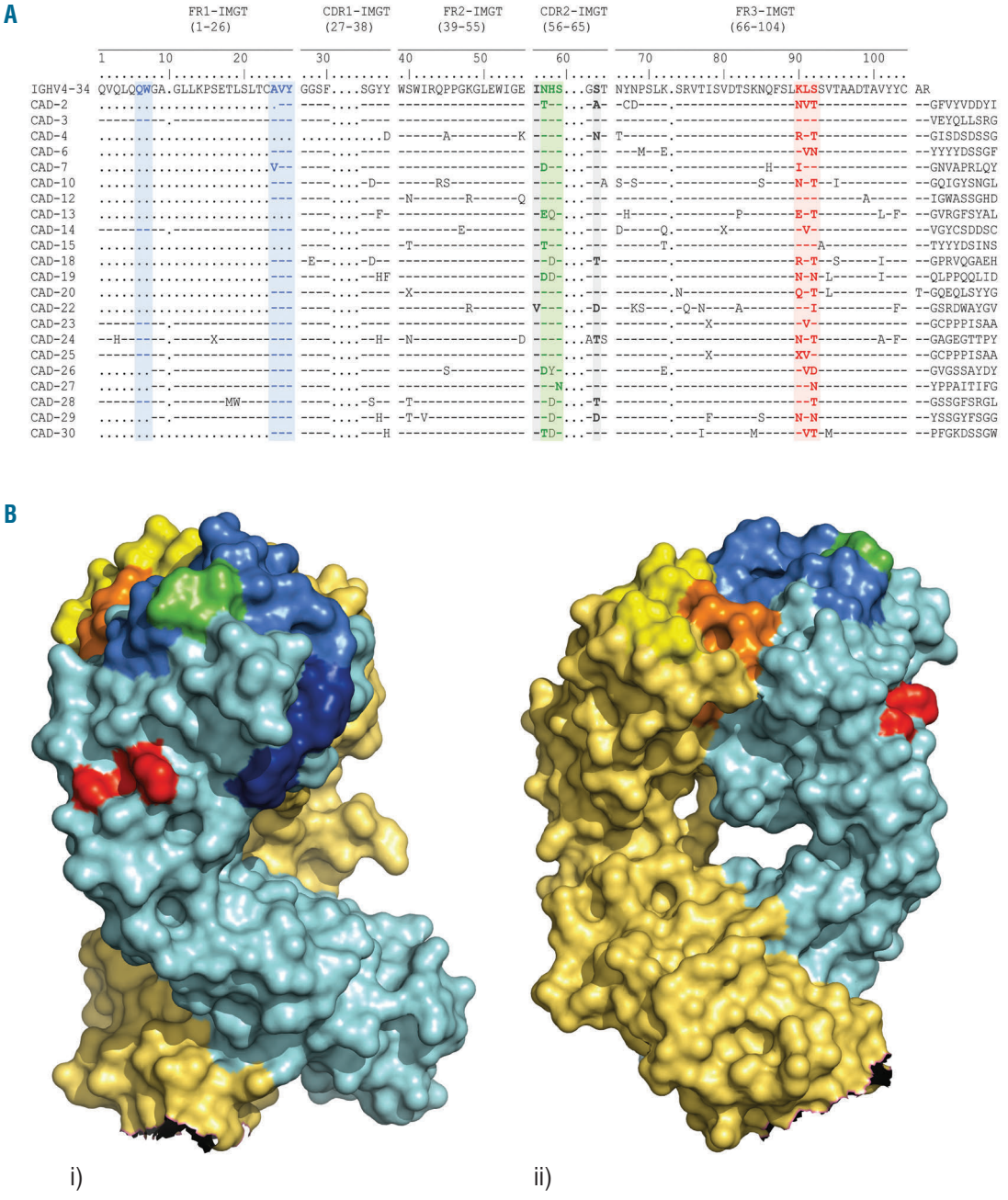

Figure 1. Molecular features of the *IGHV4-34* gene and protein in CAD patients. (A) Alignment of amino acid sequences of IGH chain from CAD patients expressing the *IGHV4-34* gene. Framework regions (FR) and complementarity determining regions (CDR) are indicated. Amino acid numbers are according to the IMGT database. Identical amino acids to the germline *IGHV4-34* sequence are indicated with a dash. For consensus numbering of amino acids according to IMGT, dots replace the missing sequence. The hydrophobic patch, required for I antigen-binding is marked in blue. The core region of the N-glycosylation site (Asn–X–Ser/Thr) is marked in green whereas the flanking residues are in bold.[3] A mutation hotspot within FR3 is marked in red. (B) Schematic representation of the crystal structure of the FAB fragment from a human IG cold agglutinin. The three-dimensional immunoglobulin structure was generated by the PyMOL program (PDB 1DN0).[8] *IGHV4-34* encoded heavy chains are colored light blue whereas *IGKV3-20* encoded light chains are colored yellow. Regions of interest are marked with different colors (coloring corresponds to A): hydrophobic patch QW + AVY (dark blue), N-glycosylation site NHS (green), KLS sequence (red), IGK CDR3 (orange), IGK CDR1 and CDR2 (bright yellow), IGH CDR1, CDR2 and CDR3 (blue) (i) View of protein surface with hydrophobic patch QW + AVY (dark blue), N-glycosylation site NHS (green) and KLS sequence (red) clearly visible. (ii) View of protein surface with IGH and IGK CDR regions visible.

tests.[2,4,5] A recent study also showed that the reintroduction of the N-glycosylation site in mutated *IGHV4-34* caused a moderate (20%), but reproducible, impairment of B cell survival in cells that are dependent on the binding of the B-cell receptor to auto-antigens.[6] SHM of the site, rendering it non-functional, might thus be advantageous for cell survival in neoplastic B cells that have escaped the normal immune control of autoreactivity. SHM of the N-glycosylation site in *IGHV4-34* in normal memory B cells may play a similar role, although in those cells increased accessibility of the antigen-binding site to foreign antigens instead of I antigen is likely important for cell selection and survival.[3]

We found that the KLS amino acid sequence in FR3 of *IGHV4-34* was mutated in almost all CAD patients (Figure 1A), and that an increased number of mutations significantly correlated with reduced hemoglobin levels at diagnosis (P=0.030) (Table 1). We also investigated mutations in this sequence in previously published sequences as well as our own database and found it often mutated, most frequently at Ser (S), in normal B cells as well as in B-cell lymphoma.[3,7] This region seems to be highly conserved in *IGHV4* family germline genes (*Online Supplementary Figure S1*). Analysis of the protein structure[8] shows that KLS amino acids are located at the protein surface, relatively close to the antigen-binding region at a distance similar to the I antigen-binding AVY hydrophobic patch (Figure 1B). The KLS amino acid sequence is coded by a DNA sequence containing five known mutation hotspot motifs characteristic for SHM (*Online Supplementary Figure S2*).[9] Clinical data suggest that a mutated KLS sequence might enhance I antigen-binding since the number of mutations in this sequence correlated with lower hemoglobin levels at diagnosis. Of interest, combining the mutation levels in N-glycosylation and KLS sequences gave a very strong correlation with decreased hemoglobin levels at diagnosis (P=0.0055) (Table 1), stressing the importance of these sites for I antigen-binding. There is no significant correlation between the overall SHM rate of *IGHV* and hemoglobin levels (P=0.21).

Few data have been published with regard to *IGKV* gene usage in CAD. Reanalysis of published data shows that the *IGKV3-20* gene was used in most published CAD cases (12/16) (see *Online Supplementary Data*). We also reanalyzed data from Li *et al.*,[4] who studied the role

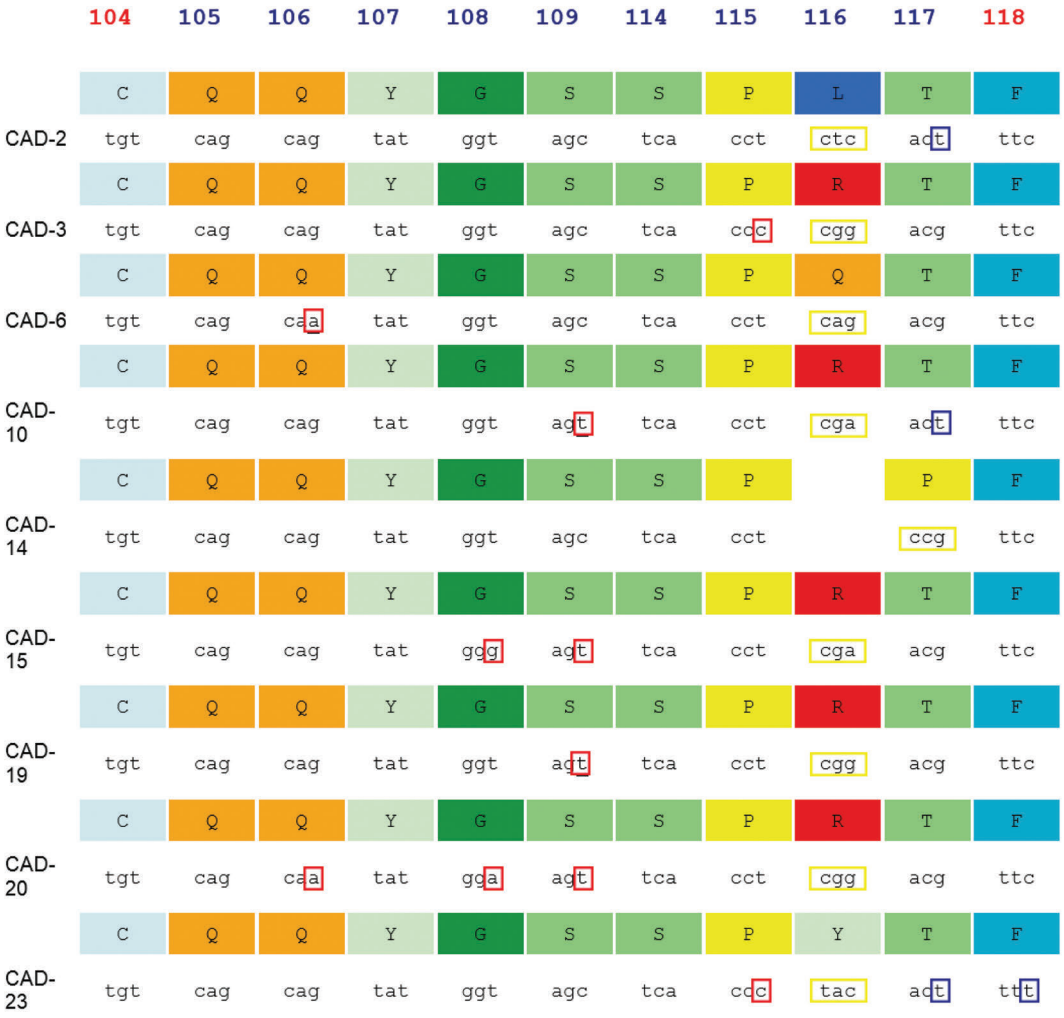

Figure 2. Alignment of highly homologous *IGKV3-20* CDR3 sequences of nine CAD patients. Somatic mutations are marked by rectangles colored red for the variable gene, blue for the joining gene and yellow for the nontemplate intervening sequence. Numbering and coloring of amino acids are according to IMGT.

of the IG light chain in determining antibody specificity by generating combinatorial antibodies (*Online Supplementary Table S2*). Antibodies with IG light chains obtained from CAD patients bound I antigen by ELISA testing, with IGKV3-20 showing the strongest binding. When combined with IGHV4-34, IG light chains obtained from other individuals without CAD showed I antigen-binding exclusively when encoded by IGKV3-20 (see *Online Supplementary Data*). Our study, including the reanalysis of older published data, shows that the IG light chain is encoded by the *IGKV3-20* or *IGKV3-15* gene in more than 80% of patients, indicating that the light chain equally contributes to I antigen-binding. The highly restricted usage of light chain genes provides a rationale for using anti-immunoglobulin light chain vaccination as part of the treatment for CAD, as has been proposed for treating other B-cell lymphomas.[10]

There was a strong correlation between younger age at diagnosis and the presence of a highly homologous or stereotyped *IGKV3-20* CDR3 region (P=0.011) (Table 1). Nine of the patients showed almost identical amino acid sequences of *IGKV3-20* CDR3 (Figure 2). This homology is a strong argument for antigen selection as a primary cause of CAD. Whether this is I antigen or a related exogenous, perhaps bacterial, antigen is unknown. Since both *IGHV4-34* and *IGKV3-20* are known to encode for antibodies that bind infectious antigens, it is possible that infection is the source of antigen selection that triggers CAD.[11] The highly homologous antibody gene sequences found in younger patients indicate that antigen selection in these patients may be different than that in other patients. Additionally, a lower mutation rate of *IGKV3-20* was also correlated with younger age at diagnosis (P=0.012) (Table 1).

Although CAD antibodies are almost exclusively encoded by *IGHV4-34*, 1 patient in the study (CAD-5) showed cold agglutinins encoded by *IGHV3-23* and *IGKV3-20*. This patient displayed typical clinical characteristics of CAD. Previously, it was shown that anti-I cold agglutinins may also be encoded by *IGHV3* family genes.[12] Marks *et al.*[13] isolated human antibody using IGHV3-23 with specificity against I antigen.

The previous finding by our group of the absence of *MYD88* L265P mutation in CAD-associated lymphoproliferative bone marrow disease[1] has recently been questioned in a review article.[14] *MYD88* L265P mutation is detected in more than 90% of lymphoplasmacytic lymphoma.[15] However, our present findings in a large number of CAD patients confirm the absence of *MYD88* L265P mutation in bone marrow samples using a sensitive technique. The absence of the mutation in primary CAD is a strong argument against the diagnosis of lymphoplasmacytic lymphoma in these patients.

In conclusion, our data show that in addition to *IGHV*, immunoglobulin light chain *V* usage is highly restricted in CAD. Our data indicate that multiple regions within the immunoglobulin heavy chain as well as the immunoglobulin light chain, contribute to I antigen-binding and may determine the clinical phenotype of the disease. Of practical consequence, the highly restricted usage of IGKV3-20 provides a rationale to include vaccination with IGKV3-20 proteins in novel clinical trial designs. IGKV3-20 proteins are known to be immunogenic and are being considered for treatment in other lymphoproliferative diseases.[10]

Agnieszka Małecka,[1] Gunhild Trøen,[1] Anne Tierens,[2] Ingunn Østlie,[1] Jędrzej Małecki,[3] Ulla Randen,[1] Sigbjørn Berentsen,[4] Geir E. Tjønnfjord,[5] and Jan M.A. Delabie[2]

[1]Department of Pathology, Oslo University Hospital, Norway; [2]Laboratory Medicine Program, University Health Network and University of Toronto, ON, Canada; [3]Department of Biological Sciences, University of Oslo, Norway; [4]Department of Research and Innovation, Haugesund Hospital, Helse Fonna, Haugesund, Norway; and [5]Department of Haematology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Norway

Correspondence: a.m.malecka@medisin.uio.no/ machynia@yahoo.com
doi:10.3324/haematol.2016.146126

Acknowledgements: the authors would like to thank the South-Eastern Norway Regional Health Authority and the Norwegian Cancer Society for supporting this study.

Key words: agglutinins, anemia, hemolysis.

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Randen U, Troen G, Tierens A, et al. Primary cold agglutinin-associated lymphoproliferative disease: a B-cell lymphoma of the bone marrow distinct from lymphoplasmacytic lymphoma. Haematologica. 2014;99(3):497-504.
2. Potter KN, Hobby P, Klijn S, Stevenson FK, Sutton BJ. Evidence for involvement of a hydrophobic patch in framework region 1 of human V4-34-encoded Igs in recognition of the red blood cell I antigen. J Immunol. 2002;169(7):3777-3782.
3. Sabouri Z, Schofield P, Horikawa K, et al. Redemption of autoantibodies on anergic B cells by variable-region glycosylation and mutation away from self-reactivity. Proc Natl Acad Sci USA. 2014; 111(25):E2567-2575.
4. Li Y, Spellerberg MB, Stevenson FK, Capra JD, Potter KN. The I binding specificity of human VH 4-34 (VH 4-21) encoded antibodies is determined by both VH framework region 1 and complementarity determining region 3. J Mol Biol. 1996;256(3):577-589.
5. Potter KN. Molecular characterization of cold agglutinins. Transfus Sci. 2000;22(1-2):113-119.
6. Young RM, Wu T, Schmitz R, et al. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proc Natl Acad Sci U S A. 2015;112(44):13447-13454.
7. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood. 2002;99(7):2562-2568.
8. Cauerhff A, Braden BC, Carvalho JG, et al. Three-dimensional structure of the Fab from a human IgM cold agglutinin. J Immunol. 2000; 165(11):6422-6428.
9. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. Nat Immunol. 2001;2(6):530-536.
10. Martorelli D, Guidoboni M, De Re V, et al. IGKV3 proteins as candidate "off-the-shelf" vaccines for kappa-light chain-restricted B-cell non-Hodgkin lymphomas. Clin Cancer Res. 2012;18(15):4080-4091.
11. Henry Dunand CJ, Wilson PC. Restricted, canonical, stereotyped and convergent immunoglobulin responses. Philos Trans R Soc Lond B Biol Sci. 2015;370(1676).
12. Jefferies LC, Carchidi CM, Silberstein LE. Naturally occurring anti-i/I cold agglutinins may be encoded by different VH3 genes as well as the VH4.21 gene segment. J Clin Invest. 1993;92(6):2821-2833.
13. Marks JD, Ouwehand WH, Bye JM, et al. Human Antibody Fragments Specific for Human Blood Group Antigens from a Phage Display Library. Biotechnology (NY) 1993;11(10):1145-1149.
14. Swerdlow SH, Kuzu I, Dogan A, et al. The many faces of small B cell lymphomas with plasmacytic differentiation and the contribution of MYD88 testing. Virchows Arch. 2016;468(3):259-275.
15. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. N Engl J Med. 2012;367(9):826-833.